

Exercise 1: Gradient Descent

Epoch = 1 :

- $f(w_1, w_2) = 0.1w_1^2 + 2w_2^2$
- $W1 = w_1 - a \cdot dw_1 = -5 - \{0.4 \cdot [0.1w_1^2]' \cdot (-5)\}$
- $= -5 - [0.4 \cdot -1] = -4.6$
- $W2 = -2 - \{0.4 \cdot [2w_2]' \cdot -2\} = -2 - 0.4 \cdot (-8) = 1.2$

Epoch = 2:

- $dw_1 = 0.2 \cdot (-4.6) = -0.92$
- $dw_2 = 4 \cdot 1.2 = 4.8$
- $w_1 = -4.6 - 0.4 \cdot -0.92 = -4.2320$
- $w_2 = 1.2 - 0.4 \cdot 4.8 = -0.7200$

Exercise 2: GD + Momentum

Epoch 1 :

- $dw_1, dw_2 = -1, -8$
- $v_1 = \text{beta} \cdot v_1 + (1 - \text{beta}) \cdot dw_1 = 0.5 \cdot 0 + (1 - 0.5) \cdot -1 = -0.5$
- $v_2 = 0.5 \cdot 0 + (1 - 0.5) \cdot -8 = -4$
- $w_1 = -5 - 0.6 \cdot -0.5 = -4.7$
- $w_2 = -2 - 0.6 \cdot -4 = 0.4$

Epoch 2 :

- $dw_1 = 0.2 \cdot -4.7 = -0.94$
- $dw_2 = 4 \cdot 0.4 = 1.6$
- $v_1 = 0.5 \cdot -0.5 + (1 - 0.5) \cdot -0.94 = -0.72$
- $v_2 = 0.5 \cdot -4 + (1 - 0.5) \cdot 1.6 = -1.2$
- $w_1 = -4.7 - 0.6 \cdot -0.72 = -4.268$
- $w_2 = 0.4 - 0.6 \cdot -1.2 = 1.12$

Exercise 3: RMSProp

Epoch 1 :

- $dw_1, dw_2 = -1, -8$
- $s_1 = \text{gamma} \cdot s_1 + (1 - \text{gamma}) \cdot dw_1^2$
- $= 0.9 \cdot 0 + (1 - 0.9) \cdot (-1)^2 = 0.1$
- $s_2 = 0.9 \cdot 0 + (1 - 0.9) \cdot (-8)^2 = 6.4$
- $w_1 = w_1 - \text{lr} \cdot (dw_1 / \sqrt{s_1 + \text{epsilon}}) = -4.051$
- $w_2 = -2 - 0.3 \cdot (-8 / \sqrt{6.4 + 10e^{-6}}) = -1.051$

Epoch 2 :

- $dw_1 = 0.2 \cdot (-4.051) = -0.8102$
- $dw_2 = 4 \cdot (-1.051) = -4.204$
- $s_1 = 0.9 \cdot 0.1 + (1 - 0.9) \cdot (-0.8102)^2 = 0.156$
- $s_2 = 0.9 \cdot 6.4 + (1 - 0.9) \cdot (-4.204)^2 = 7.527$
- $w_1 = -4.051 - 0.3 \cdot (-0.8102 / \sqrt{0.156 + 10e^{-6}}) = -3.436$
- $w_2 = -1.051 - 0.3 \cdot (-4.204 / \sqrt{7.527 + 10e^{-6}}) = -0.591$

Exercise 4: Adam

Epoch 1 :

- $dw_1, dw_2 = -1, -8$

- $v1 = \beta_1 * v1 + (1 - \beta_1) * dw1 = 0.9 * 0 + (1 - 0.9) * (-1) = -0.1$
- $v2 = 0.9 * 0 + (1 - 0.9) * -8 = -0.8$
- $s1 = \beta_2 * s1 + (1 - \beta_2) * dw1^2 = 0.999 * 0 + (1 - 0.999) * (-1)^2 = 0.001$
- $v_corr1 = v1 / (1 - \beta_1^t) = -0.1 / (1 - 0.9^1) = -1$
- $v_corr2 = -0.8 / (1 - 0.9^1) = -8$
- $s_corr1 = s1 / (1 - \beta_2^t) = 0.001 / (1 - 0.999^1) = 1$
- $s_corr2 = 0.064 / (1 - 0.999^1) = 64$
- $s2 = 0.999 * 0 + (1 - 0.999) * (-8)^2 = -0.064$
- $w1 = -5 - 0.2 * (-1) / \sqrt{1 + 10e^{-6}} = -4.8$
- $w2 = -2 - 0.2 * (-8) / \sqrt{64 + 10e^{-6}} = -1.8$

Epoch 2 :

- $dw1 = 0.2 * (-4.8) = -0.96$
- $dw2 = 4 * (-1.8) = -7.2$
- $v1 = 0.9 * (-0.1) + (1 - 0.9) * (-0.96) = -0.186$
- $v2 = 0.9 * (-0.8) + (1 - 0.9) * (-7.2) = -1.44$
- $s1 = 0.999 * 0.001 + (1 - 0.999) * (-0.96)^2 = 0.0019206$
- $s2 = 0.999 * 0.064 + (1 - 0.999) * (-7.2)^2 = 0.115776$
- $v_corr1 = -0.186 / (1 - 0.9^2) = -0.9789$
- $v_corr2 = -1.44 / (1 - 0.9^2) = -7.5789$
- $s_corr1 = 0.0019206 / (1 - 0.999^2) = 0.967078$
- $s_corr2 = 0.115776 / (1 - 0.999^2) = 57.9169$
- $w1 = -4.8 - 0.2 * (-0.9789) / \sqrt{0.967078 + 10e^{-6}} = -4.6002$
- $w2 = -1.8 - 0.2 * (-7.5789) / \sqrt{57.9169 + 10e^{-6}} = -1.6008$